A Structural Model for the Coevolution of Networks and Behavior

Konig, Michael; Hsieh, C.S.; Liu, X.

published in
Review of Economics and Statistics
2022

DOI (link to publisher)
10.1162/rest_a_00958

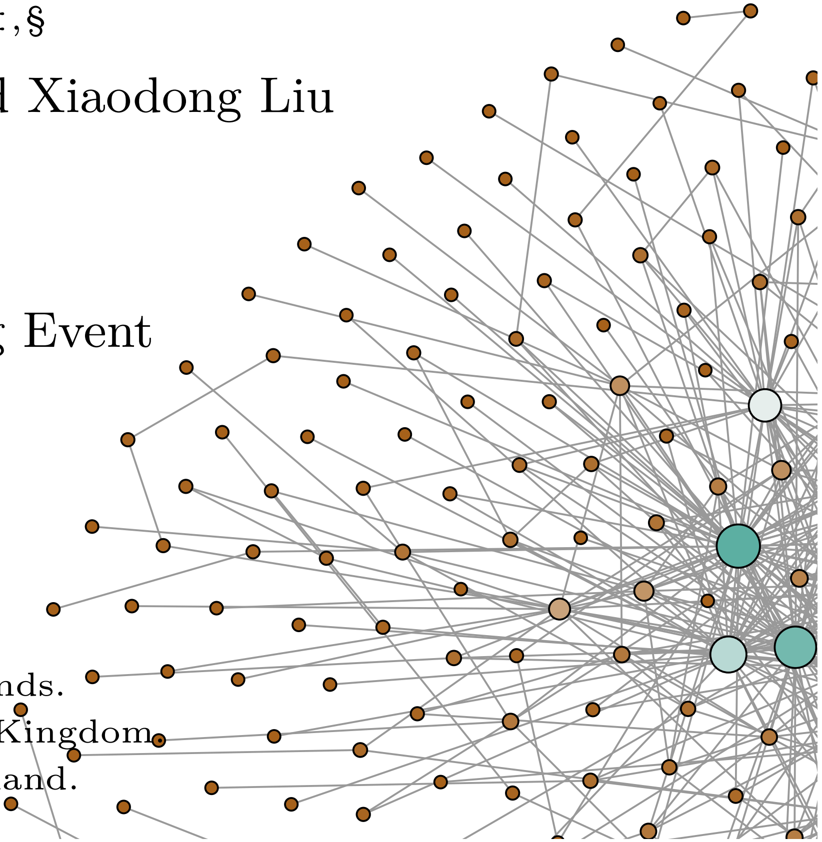
document version
Other version

A Structural Model for the Coevolution of Networks and Behavior

Michael D. König[†,‡,§]
joint with Chih-Sheng Hsieh and Xiaodong Liu

Networks Match Making Event
Kaapdoorn

20 January 2020



[†]Tinbergen Institute and VU Amsterdam, The Netherlands.
[‡]Centre for Economic Policy Research (CEPR), United Kingdom.
[§]ETH Zurich, Swiss Economic Institute (KOF), Switzerland.

Contribution

- We introduce a structural model for the *coevolution* of networks and behavior.
- The microfoundation is a network game (Ballester et al., 2006), where agents adjust actions and links in a stochastic best-response dynamics with both *strategic externalities* and *unobserved heterogeneity*.
- The coevolution process converges to a unique stationary distribution characterized by a *Gibbs measure*.
- To bypass the evaluation of the intractable normalizing constant in the Gibbs measure, we use a *Double Metropolis-Hastings algorithm* to sample from the posterior distribution of the structural parameters.
- The empirical relevance is illustrated with an example of *R&D collaborations* in the pharmaceutical industry, and we provide a long-run *key player* analysis.[1]

[1] European leaders like Emmanuel Macron and Angela Merkel have been pushing for the government supported creation and protection of European technology champions that can better compete with foreign rivals (e.g. Airbus; Source: CNN Business, 16.1.20).

Network

- Consider a *network* $G \in \mathcal{G}(n)$ consisting a set of agents $\mathcal{N} \equiv \{1, \ldots, n\}$, where $\mathcal{G}(n)$ is the set of all networks with n nodes.
- The topology of the network is represented by an $n \times n$ adjacency matrix $\mathbf{A} = [a_{ij}]$, where $a_{ij} = 1$ if agents i and j form a link and $a_{ij} = 0$ otherwise.
- The network links are reciprocal, that is, $a_{ij} = 1$ implies $a_{ji} = 1$. As a normalization, we set $a_{ii} = 0$ for all $i \in \mathcal{N}$.
- Let $\mathcal{N}_i \equiv \{j \in \mathcal{N} | a_{ij} = 1\}$ denote the set of agent i's peers (or, loosely speaking, "friends").
- Agent i, with his exogenous *characteristics* given by a (row) vector X_i, makes decisions on network links a_{ij} and the effort level y_i in an activity to maximize utility.
- We assume X_i can be observed by all the agents.

Utility

- To introduce unobserved heterogeneity in the econometric model, we allow some components of X_i to be unobservable to the econometrician.
- Let $Y = (y_1, \cdots, y_n)^\top$, and let Y_{-i} denote the effort levels of all agents but i.
- The *utility* of agent i follows a linear-quadratic function given by

$$U_i(G, Y, X) = \kappa_i(G, X) + b(X_i)y_i + \lambda \sum_{j\in\mathcal{N}} a_{ij}y_iy_j - \frac{1}{2}y_i^2, \tag{1}$$

where

$$\kappa_i(G, X) = \sum_{j\in\mathcal{N}} a_{ij}\left(\delta_0 + h(X_i, X_j, \delta_1) + \delta_2 d_{ij} + \delta_3 d_{ij}^2 + \delta_4 \sum_{k\in\mathcal{N}\setminus\{i,j\}} a_{ik}a_{jk}\right). \tag{2}$$

- The first term of Equation (1), $\kappa_i(G, X)$, captures the direct utility from the network:
  - δ_0 is the fixed *cost* of maintaining links,
  - $h(X_i, X_j, \delta_1)$ captures the similarity between agents i and j in exogenous characteristics, with the coefficient vector δ_1 representing the *homophily* effect.
  - $d_{ij} = \sum_{k \in \mathcal{N} \setminus \{i,j\}} (a_{ik} + a_{jk})$ is the total number of links of agents i and j excluding the link a_{ij}. If $\delta_2 > 0$ and $\delta_3 < 0$, then the coefficients δ_2 and δ_3 can be interpreted as the *popularity* and *congestion* effects, respectively.[2]
  - $\sum_{k \in \mathcal{N} \setminus \{i,j\}} a_{ik} a_{jk}$ is the number of common "friends" between agents i and j, with the coefficient δ_4 representing the *cyclic triangle* effect.

[2] That is, an agent may benefit from indirect links via linking to an agent with some "friends" (i.e., the popularity effect) but the marginal utility decreases as the number of indirect links increases (i.e., the congestion effect).

- **Assumption:**

$$h(X_i, X_j, \delta_1) = h(X_j, X_i, \delta_1), \ \forall i, j \in \mathcal{N}.$$

- The second term of Equation (1),

$$b(X_i) y_i,$$

  represents the *direct utility* from the effort, with the marginal utility of effort given by $b(X_i)$.
- The third term,

$$\lambda \sum_{j \in \mathcal{N}} a_{ij} y_i y_j,$$

  characterizes the social utility from the effort, with the *peer effect* given by the coefficient λ.
- Finally, we assume the *cost* of exerting effort is given by the last term of Equation (1),

$$\frac{1}{2} y_i^2,$$

  which exhibits increasing marginal cost.

Best Response

- Maximizing Equation (1) with respect to y_i gives the *best response* function for the effort choice

$$y_i = \lambda \sum_{j \in \mathcal{N}} a_{ij} y_j + b(X_i), \tag{3}$$

which coincides with the one in Ballester et al. (2006) and the Spatial Autoregressive (SAR) model.

Potential Function

- **Proposition:** The utility function defined in Equation (1) admits a potential function

$$\Phi(G, Y, X) = \kappa(G, X) + \sum_{i \in \mathcal{N}} b(X_i) y_i + \frac{\lambda}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} a_{ij} y_i y_j - \frac{1}{2} \sum_{i \in \mathcal{N}} y_i^2, \quad (4)$$

where

$$\kappa(G, X) = \frac{1}{2} \sum_{i,j \in \mathcal{N}} a_{ij} \left(\delta_0 + h(X_i, X_j, \delta_1) + \delta_2 d_{ij} + \delta_3 d_{ij}^2 + \frac{1}{3} \delta_4 \sum_{k \in \mathcal{N} \setminus \{i,j\}} a_{ik} a_{jk} \right).$$

- The potential function has the property that the change in the utility of an agent from adjusting a link or effort level is identical to the corresponding change in the potential function.

Coevolution of Networks and Behavior

- Let the realization of the network in period t be denoted by G_t with the adjacency matrix $\mathbf{A}_t = [a_{ij,t}]$, and let the network including all the current links but $a_{ij,t}$ be denoted by $G_{-ij,t}$.
- Similarly, the effort profile of $\mathcal{N}$ in period t is given by the vector $Y_t = [y_{i,t}]$, and the effort profile of $\mathcal{N}\backslash\{i\}$ is written as $Y_{-i,t}$.[3]
- The coevolution of networks and behavior is specified as a stochastic *best-response dynamics* (Blume, 1993).
- We assume time is discrete: Each time period is either
    - a link-adjustment period (with probability $0 < \rho_0 < 1$) or
    - an effort-adjustment period (with probability $1 - \rho_0$).

[3] To simplify notation, we drop X from $U_i(G, Y, X)$ and $\Phi(G, Y, X)$ henceforth.

Link Adjustment

- In a link-adjustment period, a pair of agents (i, j) is randomly selected from the population with probability $\rho(G_{t-1}, X_i, X_j)$.
- **Assumption:**
  (i) $\rho(G_{t-1}, X_i, X_j) = \rho(G_{t-1}, X_j, X_i)$;
  (ii) $\rho(G_{t-1}, X_i, X_j)$ does not depend on $a_{ij,t-1}$; and
  (iii) $\rho(G_{t-1}, X_i, X_j) > 0$ for all $(i, j) \in \mathcal{N} \times \mathcal{N}$.
- Conditional on the meeting of (i, j), agents i and j update the link a_{ij} to maximize their current utilities taking the rest of the network and effort choices as given.
- We assume that agents do not take into account the effect of their decisions on the future effort choices and network evolution.

- To capture the uncertainty (from the perspective of the econometrician) in the link adjustment process, we introduce an idiosyncratic shock to the utility.
- We assume that $a_{ij,t} = 1$ if and only if both agents i and j find that $a_{ij,t} = 1$ improves their utility, i.e.,

$$U_i(a_{ij,t} = 1, G_{-ij,t-1}, Y_{t-1}) + \epsilon_{1t} \geq U_i(a_{ij,t} = 0, G_{-ij,t-1}, Y_{t-1}) + \epsilon_{0t},$$

and

$$U_j(a_{ij,t} = 1, G_{-ij,t-1}, Y_{t-1}) + \epsilon_{1t} \geq U_j(a_{ij,t} = 0, G_{-ij,t-1}, Y_{t-1}) + \epsilon_{0t}.$$

- As

$$\begin{aligned} & \Phi(a_{ij,t} = 1, G_{-ij,t-1}, Y_{t-1}) - \Phi(a_{ij,t} = 0, G_{-ij,t-1}, Y_{t-1}) \\ = \quad & U_i(a_{ij,t} = 1, G_{-ij,t-1}, Y_{t-1}) - U_i(a_{ij,t} = 0, G_{-ij,t-1}, Y_{t-1}) \\ = \quad & U_j(a_{ij,t} = 1, G_{-ij,t-1}, Y_{t-1}) - U_j(a_{ij,t} = 0, G_{-ij,t-1}, Y_{t-1}). \end{aligned}$$

- The above two inequalities can be written more compactly as

$$\Phi(a_{ij,t} = 1, G_{-ij,t-1}, Y_{t-1}) + \epsilon_{1t} \geq \Phi(a_{ij,t} = 0, G_{-ij,t-1}, Y_{t-1}) + \epsilon_{0t}.$$

- We assume that the shocks ϵ_{0t} and ϵ_{1t} are independent from each other, i.i.d. across links and time periods and follow a Gumbel distribution.[4]
- Conditional on the meeting of (i, j), the probability that agents i and j form a link is given by

$$\mathbb{P}(a_{ij,t} = 1 | G_{-ij,t} = a_{-ij,t-1}, Y_t = Y_{t-1}) = \frac{\exp[\sigma^{-2}\Phi(a_{ij,t} = 1, G_{-ij,t-1}, Y_{t-1})]}{\exp[\sigma^{-2}\Phi(a_{ij,t} = 0, G_{-ij,t-1}, Y_{t-1})] + \exp[\sigma^{-2}\Phi(a_{ij,t} = 1, G_{-ij,t-1}, Y_{t-1})]}, \tag{5}$$

where the parameter σ^2 captures the level of "noise" in link adjustment decisions.

[4] With the distribution function $F(\epsilon) = \exp[-\exp(-\epsilon/\sigma^2)]$.

Effort Adjustment

- In an effort-adjustment period, an agent i is randomly selected from the population with probability $\rho(X_i)$.
- We assume any agent can be selected with positive probability in the following assumption.
- **Assumption:** $\rho(X_i) > 0$ for all $i \in \mathcal{N}$.
- Conditional on being selected, agent i updates the effort level $y_{it} \in \mathcal{Y}$ to maximize his current utility, where $\mathcal{Y}$ is the set of all possible effort choices.

- We allow $\mathcal{Y}$ to be continuous and assume that, taking the network G_{t-1} and the effort levels of the other agents $Y_{-i,t-1}$ as given, the probability that agent i chooses an effort level in $\mathcal{Z} \subset \mathcal{Y}$ in period t is given by

$$\mathbb{P}(y_{it} \in \mathcal{Z} | G_t = G_{t-1}, Y_{-i,t} = Y_{-i,t-1}) = \frac{\int_{\mathcal{Z}} \exp[\sigma^{-2} U_i(G_{t-1}, z, Y_{-i,t-1})] dz}{\int_{\mathcal{Y}} \exp[\sigma^{-2} U_i(G_{t-1}, y, Y_{-i,t-1})] dy}. \tag{6}$$

- Similar to Equation (5) in the link adjustment period, the probability given in Equation (6) can be justified by an additive *random utility* model over a nonfinite choice set (McFadden, 1976).
- The parameter σ^2 captures the level of "noise" in effort adjustment decisions.
- Equation (6) admits the probability density function

$$p(y_{it}|g_t = g_{t-1}, Y_{-i,t} = Y_{-i,t-1}) \\ = \frac{\exp[\sigma^{-2} U_i(g_{t-1}, y_{it}, Y_{-i,t-1})]}{\int_{\mathcal{Y}} \exp[\sigma^{-2} U_i(g_{t-1}, y, Y_{-i,t-1})] \, dy}. \quad (7)$$

Equilibrium

- The coevolution of the network G_t and effort choices Y_t follows a Markov chain, which converges to a *unique stationary distribution*.
- Let γ denote the vector of all unknown parameters in the potential function and $\theta = (\gamma^\top, \sigma^2)^\top$.
- **Proposition:** The coevolution process of the network and behavior converges to a unique stationary distribution characterized by the Gibbs measure

$$\pi(G, Y|\theta) = c(\theta)^{-1} \exp[\sigma^{-2}\Phi(G, Y|\gamma)], \tag{8}$$

where

$$c(\theta) = \sum_{G \in \mathcal{G}(n)} \int_{\mathcal{Y}^n} \exp[\sigma^{-2}\Phi(G, Y|\gamma)] \, dY.$$

Computational Problem

- Given an observation (G, Y) from the stationary distribution defined in Equation (8), we can estimate the parameter vector θ based on the maximum likelihood principle.
- However, the frequentist maximum likelihood method is impractical due to the computational difficulty in evaluating the normalizing constant

$$c(\theta) = \sum_{G \in \mathcal{G}(n)} \int_{\mathcal{Y}^n} \exp[\sigma^{-2} \Phi(G, Y|\gamma)] \, dY$$

in Equation (8):

$$\pi(G, Y|\theta) = c(\theta)^{-1} \exp[\sigma^{-2} \Phi(G, Y|\gamma)].$$

Metropolis-Hastings Algorithm

- To sample from the posterior using Markov Chain Monte Carlo (MCMC) simulation, a standard Metropolis-Hastings (MH) algorithm updates θ to $\widetilde{\theta}$, a random draw from the proposal distribution $q_\theta(\widetilde{\theta}|\theta)$, according to the acceptance probability

$$\begin{aligned}\alpha_{\theta,MH} &= \min\left\{1, \frac{p(\widetilde{\theta}|G, Y)\, q_\theta(\theta|\widetilde{\theta})}{p(\theta|G, Y)\, q_\theta(\widetilde{\theta}|\theta)}\right\} \\ &= \min\left\{1, \frac{c(\theta)\exp[\widetilde{\sigma}^{-2}\Phi(G, Y|\widetilde{\gamma})]p(\widetilde{\theta})\, q_\theta(\theta|\widetilde{\theta})}{c(\widetilde{\theta})\exp[\sigma^{-2}\Phi(G, Y|\gamma)]p(\theta)\, q_\theta(\widetilde{\theta}|\theta)}\right\}.\end{aligned}$$

- The computational problem still exists as $c(\theta)$ and $c(\widetilde{\theta})$ in the acceptance probability do not cancel each other.

Exchange Algorithm

- A way to bypass the evaluation of the intractable normalizing constant $c(\theta)$ is to use the *exchange algorithm*, which takes the following steps at each iteration:

  Step 1 Draw $\widetilde{\theta}$ from the proposal distribution $q_\theta(\widetilde{\theta}|\theta)$.

  Step 2 Generate $(\widetilde{G}, \widetilde{Y})$ from the distribution $\pi(G, Y|\widetilde{\theta})$ using a perfect sampler.

  Step 3 Accept $\widetilde{\theta}$ according to the acceptance probability

$$\begin{aligned}\alpha_{\theta,EX} &= \min\left\{1, \frac{p(\widetilde{\theta}|G, Y)\,q_\theta(\theta|\widetilde{\theta})\pi(\widetilde{G}, \widetilde{Y}|\theta)}{p(\theta|G, Y)\,q_\theta(\widetilde{\theta}|\theta)\pi(\widetilde{G}, \widetilde{Y}|\widetilde{\theta})}\right\} \\ &= \min\left\{1, \frac{\exp[\widetilde{\sigma}^{-2}\Phi(G, Y|\widetilde{\gamma})]p(\widetilde{\theta})\,q_\theta(\theta|\widetilde{\theta})\exp[\sigma^{-2}\Phi(\widetilde{G}, \widetilde{Y}|\gamma)]}{\exp[\sigma^{-2}\Phi(G, Y|\gamma)]p(\theta)\,q_\theta(\widetilde{\theta}|\theta)\exp[\widetilde{\sigma}^{-2}\Phi(\widetilde{G}, \widetilde{Y}|\widetilde{\gamma})]}\right\}.\end{aligned}$$

- The main advantage of the exchange algorithm is that the acceptance probability does not contain the normalizing constant $c(\theta)$ and thus can be evaluated.

Double Metropolis-Hastings Algorithm

- In Step 2 of the exchange algorithm, we need to generate auxiliary date using a perfect sampler.
- This is computationally costly for our model and, more generally, exponential random graph models (ERGMs).
- To overcome this issue, Liang (2010) proposed a DMH algorithm, which uses a finite run of the MH algorithm initialized at the observed (G, Y) to generate auxiliary data $(\widetilde{G}, \widetilde{Y})$.

- More specifically, at each iteration, the DMH algorithm follows the same steps as the exchange algorithm with the second step replaced by:[5]

  Step 2* Generate $(\widetilde{G}, \widetilde{Y})$ from the distribution $\pi(G, Y|\widetilde{\theta})$ using a finite run of the MH algorithm initialized at the observed (G, Y).

- To generate auxiliary data $(\widetilde{G}, \widetilde{Y})$, one could use a local sampler. However, the convergence of such a sampler could be slow in practice.
- To improve convergence and reduce computational burden, we propose an MH algorithm to generate $(\widetilde{G}, \widetilde{Y})$.

[5] Compared with Mele (2017), one additional complication is that we need to simulate both networks $\widetilde{G}$ and effort choices $\widetilde{Y}$ in Step 2* of the DMH algorithm.

Algorithm: Auxiliary Data Generation

Given θ, at each iteration:

Step 1 Draw $\widetilde{G}$ from the proposal distribution $q_G(\widetilde{G}|G)$. Let $\widetilde{A}$ denote the adjacency matrix of $\widetilde{G}$.

Step 2 Generate $\widetilde{Y} \sim \mathcal{N}(\widetilde{Y}^*, \Sigma_{\widetilde{Y}})$, where $\widetilde{Y}^* \equiv (I_n - \lambda\widetilde{A})^{-1}B(X)$, with $B(X) = [b(X_1), \cdots, b(X_n)]^\top$, is the equilibrium effort vector derived from the best response function (3), and $\Sigma_{\widetilde{Y}} = \sigma^2(I_n - \lambda\widetilde{A})^{-1}$.

Step 3 Accept $(\widetilde{G}, \widetilde{Y})$ according to the acceptance probability

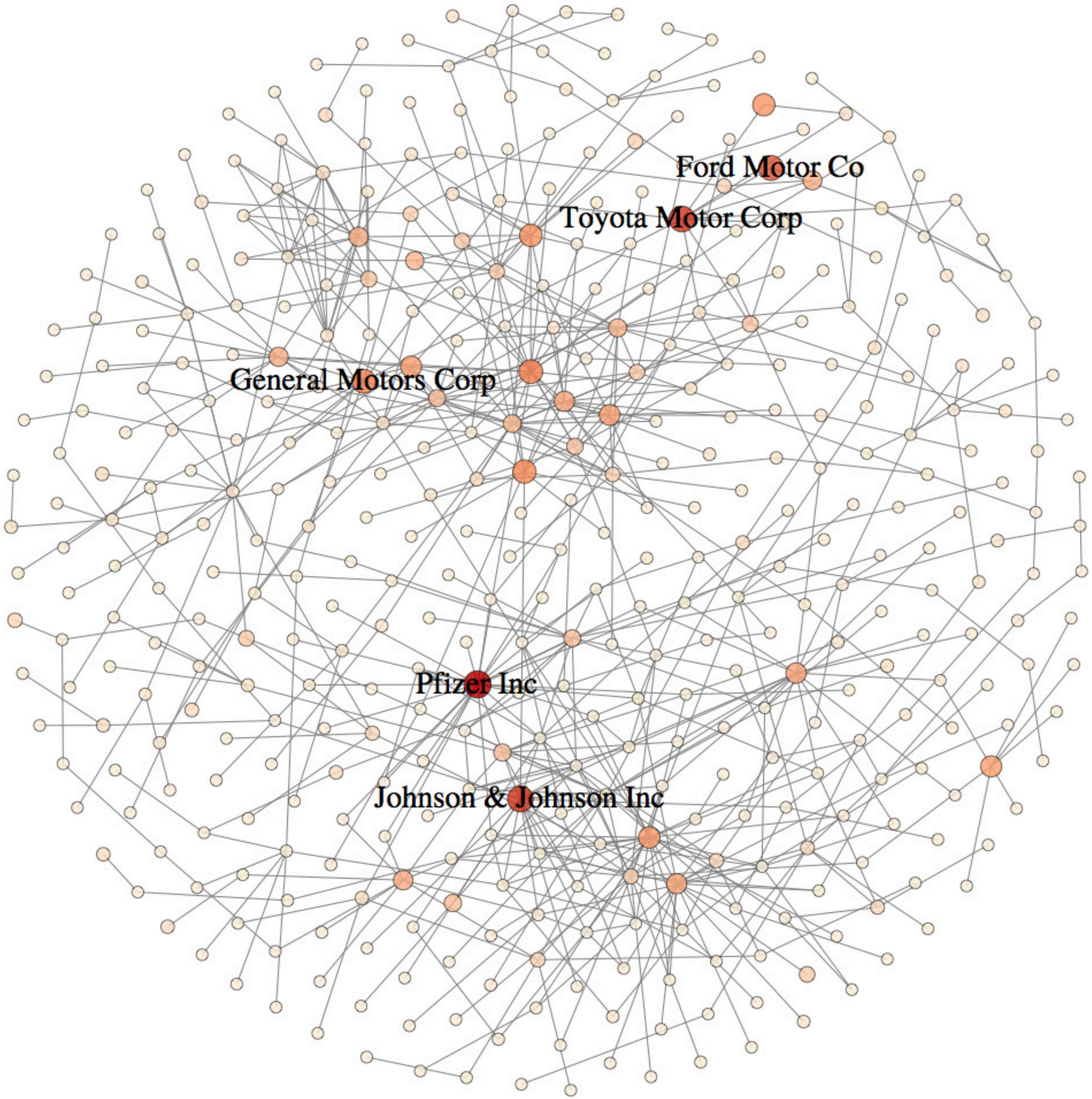
$$\begin{aligned}\alpha_{(G,Y),MH} &= \min\left\{1, \frac{\pi(\widetilde{G}, \widetilde{Y}|\theta)p_Y(Y|G)q_G(G|\widetilde{G})}{\pi(G, Y|\theta)p_Y(\widetilde{Y}|\widetilde{G})q_G(\widetilde{G}|G)}\right\} \\ &= \min\left\{1, \frac{\exp[\sigma^{-2}\Phi(\widetilde{G}, \widetilde{Y}|\gamma)]p_Y(Y|G)q_G(G|\widetilde{G})}{\exp[\sigma^{-2}\Phi(G, Y|\gamma)]p_Y(\widetilde{Y}|\widetilde{G})q_G(\widetilde{G}|G)}\right\},\end{aligned}$$

where $p_Y(\widetilde{Y}|\widetilde{G})$ denotes the density function of $\mathcal{N}(\widetilde{Y}^*, \Sigma_{\widetilde{Y}})$.

R&D Networks

- We merged the MERIT-CATI with the Thomson SDC *alliance databases*.[6]
- We use annual data about *balance sheets* and income statements from Standard & Poor's Compustat and Bureau Van Deijk's Orbis databases.
- We also obtained the firms' *patents* (PATSTAT), and computed the potential technology spillovers between collaborating firms using various patent proximity indices.

[6] These databases contain information about strategic technology agreements, including any alliance that involves some arrangements for mutual transfer of technology or joint research, such as joint research pacts, joint development agreements, cross licensing, R&D contracts, joint ventures and research corporations. Cf. Schilling (SMJ, 2009) and Hagedoorn (RP, 2002).


Ford Motor Co
Toyota Motor Corp
General Motors Corp
Pfizer Inc
Johnson & Johnson Inc

Estimation Results

Table: Estimation results.

		Model 1		Model 2	
spillover effect	(λ)	0.0082	(0.0021)***	0.0060	(0.0013)***
Production Cost					
productivity	(β_1)	0.8662	(0.0192)***	0.8877	(0.0056)***
unobs. heterogeneity	(β_2)			0.0782	(0.0468)
sub-sector dummies		Yes		Yes	
Collaboration Cost					
constant	(δ_0)	-4.7166	(0.0917)***	-5.1451	(0.1782)***
same sub-sector	(δ_{11})	0.5322	(0.0275)***	0.8163	(0.1299)***
diff-in-productivity	(δ_{12})	0.0708	(0.0192)***	0.0395	(0.0225)*
popularity	(δ_2)	0.0011	(0.0318)	0.0608	(0.0349)*
congestion	(δ_3)	-0.0036	(0.0014)**	-0.0056	(0.0015)***
cyclic triangle	(δ_4)	0.4547	(0.0535)***	0.3531	(0.1088)***
Noise Parameters					
noise in decisions	(σ^2)	0.3835	(0.0228)***	0.4262	(0.0217)***
unobs. heterogeneity	(ς_x^2)			0.8609	(0.0760)***

Notes: Model 1 ignores unobserved heterogeneity and Model 2 controls for unobserved heterogeneity. Standard errors in parentheses. ***, **, and * indicate that the highest density range does not cover zero at 99%, 95%, and 90% levels.

Identifying Key Players

	SIC	mkt. share	R&D exp.	deg.	w/o network rewiring welfare loss	w/o network rewiring rank	with network rewiring welfare loss	with network rewiring rank
Johnson & Johnson Inc.	283	3.0547	15.1535	7	-0.5237	3	-0.3511	1
Wyeth	283	1.1686	14.2487	2	-0.5373	1	-0.3219	2
Schering-Plough Corp.	283	0.6057	13.8905	1	-0.5158	4	-0.3096	3
Bristol-Myers Squibb Co.	283	1.0287	14.2351	6	-0.5255	2	-0.2924	4
Pfizer Inc.	283	2.7679	15.2467	15	-0.4712	9	-0.2843	5
Unilever PLC	284	5.4914	13.2439	0	-0.5025	7	-0.2694	6
Abbott Laboratories Inc.	283	1.2907	14.5658	3	-0.5088	5	-0.2553	7
Merck & Co Inc.	283	1.2999	14.6794	10	-0.5046	6	-0.2515	8
Akzo Nobel NV	285	11.7496	13.2205	2	-0.4450	13	-0.2423	9
Bayer	280	3.8340	14.1742	10	-0.4562	11	-0.2361	10
Daiichi Sankyo Co. Ltd.	283	0.4590	13.4980	5	-0.4239	20	-0.2354	11
Elsai	283	0.3329	13.0432	1	-0.4673	10	-0.2297	12
L'Oreal SA	284	2.1873	12.7125	0	-0.4545	12	-0.2276	13
Novartis	283	2.0691	14.7913	15	-0.4715	8	-0.2190	14
Asahi Kasei Corp.	280	1.4715	12.3177	0	-0.4370	16	-0.2171	15
Merck KGaA	283	0.4515	13.0571	3	-0.4419	14	-0.2094	16
Henkel AG & Co. KGaA	284	1.7648	12.2638	0	-0.4243	19	-0.2022	17
Solvay SA	280	1.2445	12.7682	3	-0.3612	41	-0.1948	18
Kaocorp	284	1.1679	12.1513	0	-0.4272	18	-0.1948	19
Shionogi & Co. Ltd.	283	0.0986	11.9814	0	-0.4161	22	-0.1908	20

Notes: The market share is defined as the percentage of a firm's sales in a sub-sector at the 3-digit SIC level. Welfare loss is measured in percentage.